

We claim:

1. A computer-implemented method for constructing a single vector representing a semantic abstract in a topological vector space for a semantic content of a document on a computer system, the method comprising:

5 storing a semantic content for the document in computer memory accessible by the computer system;

constructing state vectors in the topological vector space for the semantic content;

superpositioning the state vectors to construct the single vector; and

10 storing the single vector as the semantic abstract for the document.

2. A method according to claim 1, wherein constructing the state vectors includes:

identifying lexemes/lexeme phrases in the semantic content; and

15 constructing a state vector for each lexeme/lexeme phrase in the semantic content.

3. A method according to claim 1, wherein superpositioning the state vectors includes adding the state vectors using vector arithmetic.

4. A method according to claim 1, wherein superpositioning the state vectors includes weighting the state vectors.

5. A method according to claim 1 further comprising normalizing the single vector.

6. A method according to claim 1, wherein:

25 storing a semantic content includes:

storing the document in computer memory accessible by the computer system;

and

extracting words from at least a portion of the document;

30 constructing state vectors includes constructing a state vector in the topological vector space for each word using a dictionary and a basis; and

the method further comprises filtering the state vectors.

MJM Docket No. 6647-16

[illegible]

the program further comprises filtering software to filter the state vectors.

13. An apparatus on a computer system to construct a single vector representing a semantic abstract in a topological vector space for a semantic content of a document on a

5 computer system, the apparatus comprising:

a semantic content stored in a memory of the computer system;

a state vector constructor for constructing state vectors in the topological vector space for the semantic content; and

10     a superpositioning unit adapted to superposition the state vectors into a single vector  
        as the semantic abstract.

14. An apparatus according to claim 13, wherein:

the state vector includes an associated threshold distance; and

the apparatus further comprises:

15 search means for searching the topological vector space for a second document with a second semantic abstract within the threshold distance associated with the first semantic abstract for the first document; and

retrieval means to retrieve the second document.

15. An apparatus according to claim 13, wherein the state vector constructor includes a lexeme identifier adapted to identify lexemes/lexeme phrases in the semantic content.

16. An apparatus according to claim 13, wherein the superpositioning unit  
25 includes a vector arithmetic unit adapted to add the state vectors.

17. An apparatus according to claim 13 further comprising a normalization unit adapted to normalize the single vector.

30            18.    An apparatus according to claim 13, wherein:  
the apparatus further comprises:

a lexeme extractor adapted to extract lexemes/lexeme phrases from the semantic content; and

filtering means for filtering the state vectors; and  
the state vector constructor is adapted to constructing a state vector in the topological vector space for each lexeme/lexeme phrase using a dictionary and a basis.

5           19.     A computer-implemented method for constructing minimal vectors  
representing a semantic abstract in a topological vector space for a semantic content of a  
document on a computer system, the method comprising:  
              storing a semantic content for the document in computer memory accessible by the  
computer system;  
10           constructing state vectors in the topological vector space for the semantic content;  
              locating clumps of state vectors in the topological vector space;  
              superpositioning the state vectors within each clump to form a single vector  
representing the clump;  
              collecting the single vectors representing each clump to form the minimal vectors; and  
15           storing the minimal vectors as the semantic abstract for the document.

              20.     A computer-readable medium containing a program to construct minimal  
vectors representing a semantic abstract in a topological vector space for a semantic content  
of a document on a computer system, the program comprising:  
20           storing software to store a semantic content for the document in computer memory  
accessible by the computer system;  
              construction software to construct state vectors in the topological vector space for the  
semantic content;  
              clump location software to locate clumps of state vectors in the topological vector  
25           space;  
              superpositioning software to superposition the state vectors within each clump to form  
a single vector representing the clump;  
              collection software to collect the single vectors representing each clump to form the  
minimal vectors; and  
30           storing software to store the minimal vectors as the semantic abstract for the  
document.

21. An apparatus on a computer system to construct minimal vectors representing a semantic abstract in a topological vector space for a semantic content of a document on a computer system, the apparatus comprising:

a semantic content stored in a memory of the computer system;

5 a state vector constructor for constructing state vectors in the topological vector space for the semantic content;

a clump locator unit adapted to locate clumps of state vectors in the topological vector space;

10 a superpositioning unit adapted to superposition the state vectors within each clump into a single vector representing the clump; and

a collection unit adapted to collect the single vectors representing the clump into the minimal vectors of the semantic abstract.

09591520 101000  
000000 000000